

# METHOD AND DEVICE FOR SUMMARIZING VIDEO IMAGE

Publication number: JP2001119649

Publication date: 2001-04-27

Inventor: ECHIGO TOMIO; TOMITA ALBERT; MASUMITSU TAKESHI

Applicant: IBM

Classification:

- international: G06F17/30; H04N5/76; H04N5/91; G06F17/30; H04N5/76; H04N5/91; (IPC1-7): H04N5/76; G06F17/30; H04N5/91

- european:

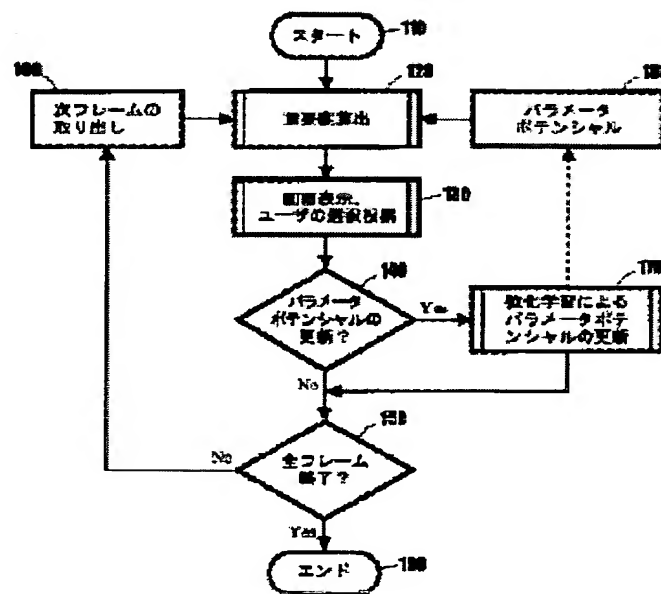
Application number: JP19990300349 19991022

Priority number(s): JP19990300349 19991022

Report a data error here

## Abstract of JP2001119649

**PROBLEM TO BE SOLVED:** To generate a summary of video images along user's preference. **SOLUTION:** This device generates a summary from video images and is provided with a means that extracts numerical data from objective video images, a means that decides an important part of each scene on the basis of the extracted numerical data and a means that generates a summary from parts of the importance, that have a value which is a prescribed threshold or higher.



全体のフローチャート

Data supplied from the esp@cenet database - Worldwide

THIS PAGE BLANK (USPTO)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-119649

(P2001-119649A)

(43) 公開日 平成13年4月27日 (2001.4.27)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テ-マコ-ド* (参考)
H 0 4 N 5/76		H 0 4 N 5/76	B 5 B 0 7 5
G 0 6 F 17/30		G 0 6 F 15/40	3 7 0 D 5 C 0 5 2
H 0 4 N 5/91		H 0 4 N 5/91	N 5 C 0 5 3

審査請求 有 請求項の数12 O L (全 13 頁)

(21) 出願番号 特願平11-300349

(22) 出願日 平成11年10月22日 (1999. 10. 22)

(71) 出願人 390009531

インターナショナル・ビジネス・マシー  
ズ・コーポレーション

INTERNATIONAL BUSIN  
ESS MACHINES CORPO  
RATION

アメリカ合衆国10504、ニューヨーク州

アーモンク (番地なし)

(74) 代理人 100086243

弁理士 坂口 博 (外1名)

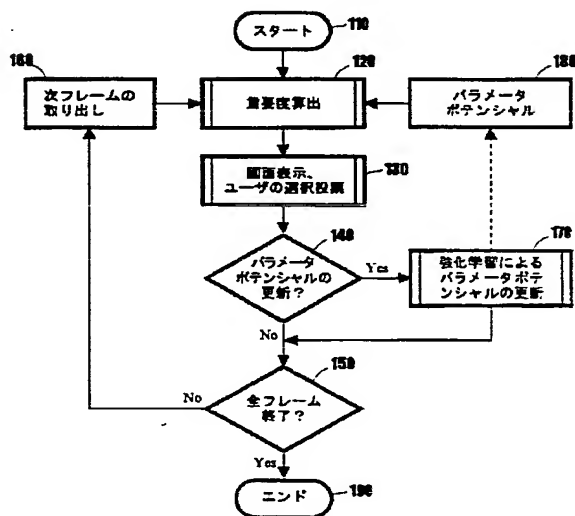
最終頁に続く

(54) 【発明の名称】 映像の要約方法及び装置

(57) 【要約】

【課題】本願発明が解決しようとする課題は、ユーザの好みに沿った映像の要約を作成することである。

【解決の手段】本発明は、映像から要約を作成する装置であって、対象となる映像から数値データを抽出する手段、前記抽出した数値データに基づいて各シーンの重要度を決定する手段と、前記重要度が所定の閾値以上の部分から要約を作成する手段を含む。



全体のフローチャート

## 【特許請求の範囲】

【請求項 1】映像から要約を作成する方法であって、  
対象となる映像から数値データを抽出するステップと、  
前記抽出した数値データに基づいて各シーンの重要度を  
決定するステップと、  
前記重要度が所定の閾値以上の部分から要約を作成する  
ステップと、  
を含む方法。

【請求項 2】映像から要約を作成する方法であって、  
対象となる映像から数値データを抽出するステップと、  
前記対象となる映像がセマンティック・データを含む場  
合には当該セマンティック・データ及び前記数値データ  
に基づいて各シーンの重要度を決定し、前記対象となる  
映像がセマンティック・データを含まない場合には、前  
記数値データに基づいて各シーンの重要度を決定するス  
テップと、  
前記重要度が所定の閾値以上の部分から要約を作成する  
ステップと、  
を含む方法。

【請求項 3】映像から要約を作成する方法であって、  
対象となる映像から複数の種類の数値データを抽出する  
ステップと、  
前記抽出した複数の種類の数値データに基づいて各シー  
ンの重要度を決定するステップと、  
前記重要度が所定の閾値以上の部分から要約を作成する  
ステップと、  
を含む方法。

【請求項 4】映像から要約を作成する方法であって、  
対象となる映像から数値データを抽出するステップと、  
対象となる映像を表示中にユーザの好みを表すデータを  
受け取るステップと、  
前記抽出した数値データ及び前記ユーザの好みを表すデ  
ータに基づいて各シーンの重要度を決定するステップ  
と、  
前記重要度が所定の閾値以上の部分から要約を作成する  
ステップと、を含む方法。

【請求項 5】映像から要約を作成する方法であって、  
対象となる映像から数値データを抽出するステップと、  
前記抽出した数値データに基づいて各シーンの重要度を  
決定するステップと、  
前記重要度が所定の閾値以上の部分から要約を作成する  
ステップと、  
前記作成された要約を表示するステップと、  
前記要約を表示中にユーザの入力により前記重要度を改  
変するステップと、  
を含む方法。

【請求項 6】映像から要約を作成する装置であって、  
対象となる映像から数値データを抽出する手段と、  
前記抽出した数値データに基づいて各シーンの重要度を  
決定する手段と、

前記重要度が所定の閾値以上の部分から要約を作成する  
手段と、  
を含む装置。

【請求項 7】映像から要約を作成する装置であって、  
対象となる映像から数値データを抽出する手段と、  
前記対象となる映像がセマンティック・データを含む場  
合には当該セマンティック・データ及び前記数値データ  
に基づいて各シーンの重要度を決定し、前記対象となる  
映像がセマンティック・データを含まない場合には、前  
記数値データに基づいて各シーンの重要度を決定する手  
段と、  
前記重要度が所定の閾値以上の部分から要約を作成する  
手段と、  
を含む方法。

【請求項 8】映像から要約を作成する装置であって、  
対象となる映像から複数の種類の数値データを抽出する  
手段と、  
前記抽出した複数の種類の数値データに基づいて各シー  
ンの重要度を決定する手段と、  
前記重要度が所定の閾値以上の部分から要約を作成する  
手段と、  
を含む装置。

【請求項 9】映像から要約を作成する装置であって、  
対象となる映像から数値データを抽出する手段と、  
対象となる映像を表示中にユーザの好みを表すデータを  
受け取る手段と、  
前記抽出した数値データ及び前記ユーザの好みを表すデ  
ータに基づいて各シーンの重要度を決定する手段と、  
前記重要度が所定の閾値以上の部分から要約を作成する  
手段と、  
を含む装置。

【請求項 10】映像から要約を作成する装置であって、  
対象となる映像から数値データを抽出する手段と、  
前記抽出した数値データに基づいて各シーンの重要度を  
決定する手段と、  
前記重要度が所定の閾値以上の部分から要約を作成する  
手段と、  
前記作成された要約を表示する手段と、  
前記要約を表示中にユーザの入力により前記重要度を改  
変する手段と、  
を含む装置。

【請求項 11】映像から要約を作成するシステムであっ  
て、  
前記システムは、  
プロセッサと、  
メモリと、  
入力された映像を記憶する記憶装置と、  
映像要約装置とを含み、  
前記映像要約装置は、  
対象となる映像から数値データを抽出する手段と、

前記抽出した数値データに基づいて各シーンの重要度を決定する手段と、  
前記重要度が所定の閾値以上の部分から要約を作成する手段と、  
を含むシステム。

【請求項 12】映像から要約を作成するプログラムを記憶する記録媒体であって、

前記プログラムは、  
対象となる映像から数値データを抽出するステップと、  
前記抽出した数値データに基づいて各シーンの重要度を決定するステップと、  
前記重要度が所定の閾値以上の部分から要約を作成するステップと、  
を含む記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、動画データを処理可能なコンピュータ・システム、特に画像処理装置における動画データの要約方法及び装置に関する。

【0002】

【従来の技術】2000年のBSデジタル・テレビ放送に続き、地上波デジタル放送が2003年に始まろうとしているが、そのときには現在のアナログ放送とは異なった視聴方法が考えられる。

【0003】従来のアナログ放送では、ユーザはリアル・タイムにライブを視聴したり、又は特定の時間にセットしたビデオ・デッキに映像を録画し、録画開始時間の頭出ししかできなかった。

【0004】本発明の対象となる映像とは、時間とともに変化する画像のことであり、動画データ、ビデオ・データ等と同義である。

【0005】しかし、デジタル放送が普及するころには、低価格で大容量のハードディスク装置(HDD)が受像機であるセットトップ・ボックス(Set-Top Box、STB)に組み込まれ、ユーザの好みによる長時間映像が録画可能になるが、ユーザの総視聴時間は、いつの時代でも大きく変わらないので、効率良く映像を検索し、要約して提示することが更に重要となる。

【0006】ここで、セットトップ・ボックス(STB)とは、ケーブルTV(CATV)の制御装置や、通信カラオケの端末機器など、テレビに接続して追加機能を提供するデバイスの一般名称であり、通常TVの上に置くことからこのように呼ばれている。

【0007】例えば、スポーツ映像を途中から見る場合、それまでの経過を要約して提示し、ライブ映像に追いつくなどの映像表現が考えられている。

【0008】しかし、従来の要約法では要約結果は、コンテンツの配信者である放送局が決めており、個人の好みに沿った要約を生成することが困難であった。

【0009】ここで要約とは、対象となる映像から時間

軸に圧縮した映像を生成することであり、長時間の対象映像からユーザの興味のないシーケンスを削除し、興味のあるシーケンスだけを残したものである。

【0010】映像の要約には、内容解析を必要とするため、高性能なパーソナル・コンピュータ(PC)を必要としていたが、今後放送が開始されるデジタル放送では、映像の隙間にデータ放送を配信することができるため、配信者が用意する複数のデータをデータ放送で送り、ユーザが自由にデータを組み合わせてユーザ個々に適した要約を生成することが可能となる。

【0011】また、画像処理機能を有するSTBで処理したデータを利用した要約も可能である。

【0012】従来からの映像の要約技術として、連続フレームの画素値がフレーム内で大きく異なる時間を見出すシーン・チェンジの先頭画像を並べたストーリー・ボード(story board)をあげている例がある。

【0013】しかし、ストーリー・ボード自体はシーンの内容を伝えていないため要約にはならない。

【0014】ストーリー・ボードの画像の類似度を求め、関係グラフで一連のシーンを集約したシーン・トランジション・グラフ(Scene Transition Graph) [1]

「M. Yeung, B.L. Yeo, etc., "Extracting story units from long programs for video browsing and navigation," IEEE ICICS, pp. 296-305, 1996.」は、映像の構造を伝えることが可能であるが、要約にはならない。

【0015】要約技術では、ビデオ・スキミング法(Video Skimming) [2] 「M. Smith and T. Kanade, "Video skimming and characterization through the combination of image and language understanding techniques," IEEE CVPR, pp. 775-781, 1997.」が映像だけでなく、音声、クローズド・キャプションを用いて、有効性を示しているが、個人の好みを反映するための学習機能はない。

【0016】米国においては、STBが普及し始めており、ハード・ディスク装置(HDD)を有する長時間録画可能な機種が発表されており、MbTV、Tivo社等が有力なメーカであり、個人の好みの映像を蓄積するアルゴリズムとして、ベイズ(Bays)法が使われている([3] 「MbTV, Preference Determination Engine」)。

【0017】しかし、これらは映像タイトル単位で蓄積する方法であり、後述する本発明のように内容を要約することは行っていない。

【0018】本発明を用いると、配信者によって送られた複数のデータ又はユーザのSTBによって処理したデータを用い、ユーザは個人の好みによって見たいシーンを重要視するデータに対する重み付けをすることができる。

【0019】映像の内容を表現するデータのタイプとして、シーンの意味を直接表現するセマンティック・デー

10

20

30

40

50

タ (semantic data) と映像における信号の強さを表す数値データ (numerical data) がある。

【0020】セマンティック・データは映像に含まれる各シーンの内容を言葉で表したものであり、数値データは、映像に含まれる各シーンの内容を数値で表現したものである。

【0021】ここでシーンとは、一般的に意味解釈を与えることができるフレーム又はフレーム列のことである。

【0022】ユーザにとって セマンティック・データは入力と結果の関係が明瞭であるが、数値データの場合はどのように作用するか分かりづらい。

【0023】セマンティック・データは映像シーンを解釈する必要があるため、セマンティック・データだけに頼った要約をするには、粒度の細かな注釈が必要となる。

【0024】通常セマンティック・データは、人手により入力されるため、粒度の細かい詳細な注釈を映像に付けるためには時間や経費がかかり、また、セマンティック・データの複数のパラメータを関連付けるためには、  
20 正確な定義文が必要となる。

【0025】全ての映像に、粒度の細かな セマンティック・データが付帯しているとは限らないので、セマンティック・データが無い場合の要約を数値データのみから行う必要がある。

【0026】どのような数値データが有効であるかは、過去に採用されたシーンから重み付けることができる。

【0027】本発明を用いると、各シーンを採用するかスキップするかで数値データのパラメータ・ポテンシャルを変更でき、ユーザ各人の映像重要度を設定することができ、期待される要約時間に近似した要約映像を得ることができる。

【0028】本発明により、長時間の映像からユーザが好むシーンを含む映像だけを取り出すことができ、わずかな視聴時間で興味ある映像を楽しむことができる。

【0029】映像に付帯するデータは、データ放送で送信できるように、映像に比べてデータ容量は小さく、安価なSTBで要約処理が可能である。

【0030】映像に付帯するデータが十分でない場合は、STBで映像を処理する必要があるが、処理済のデータを使えば、同じ構成で要約が可能となる。  
40

【0031】

【発明が解決しようとする課題】本願発明が解決しようとする課題の1つは、長時間映像をユーザの望む短時間のダイジェストに要約するための重要度が、各ユーザ毎に変更可能となる機能を提供することである。

【0032】

【課題を解決するための手段】本発明は、映像から要約を作成する装置であり、対象となる映像から数値データを抽出する手段、前記抽出した数値データに基づいて各  
50

シーンの重要度を決定する手段、前記重要度が所定の閾値以上の部分から要約を作成する手段を含む。

【0033】

【発明の実施の形態】本発明の実施例として、以下でサッカーのライブ中継の要約を紹介する。

【0034】現状では、放送局は映像に付随するセマンティック・データ等の情報を付加していないため、要約のための画像処理は、ユーザ自身が行うことになる。

【0035】ユーザが不在の間に、本発明を実装したSTBは、サッカー映像を蓄積（録画）するとともに前もって定義した画像処理を実行し（追跡、動作分類、カメラモーション検出、選手の位置推定）、画像特徴（パラメータ）を抽出する。

【0036】また同時に、以前から作成してある個人所有のパラメータ・ポテンシャルや2次元ポテンシャル・グラフを利用して、その放送の映像重要度を決定しておく。

【0037】視聴しながら興味のない要約が現れた場合は、映像を早送りしてその場面をスキップさせる。

【0038】ユーザが視聴し終わるとSTBは、その日のユーザの視聴行為を学習するため、要約が受け入れられたシーンに対応するパラメータを、パラメータ・ポテンシャルと2次元パラメータ・グラフに投票して加算し、要約をスキップしたシーンに対応するパラメータを、パラメータ・ポテンシャルと2次元ポテンシャル・グラフから減算する。

【0039】もし、放送局が映像に付随するセマンティック・データ等の意味データを送信している場合には、ユーザは要約生成の前に、見たいキーワード（選手名やプレーの種類など）を入力し、映像重要度に加えることも可能であるが必須ではない。

【0040】最初、ユーザが個人の重要度を決定するパラメータ・ポテンシャルを持っていないとき、映像配信者が送ってくる映像重要度をサンプルとして用いるか、ユーザが個人で見たいシーンのキーワードを入力して要約を生成するか、又は映像の早送りをを用いた個人の要約をサンプルとして生成する必要がある。

【0041】サンプルでは、重要度は時間毎の一次元グラフで与えられ、その値は0～1までの実数とし、高い重要度を持つ時間帯にはシーンを記述する注釈が付けられている（図1）。

【0042】次にユーザは、セマンティック・データがある場合はセマンティック・データから好みの注釈を選択することにより、サンプル映像重要度は変更され、変更した重要度から自動的に得られるダイジェスト映像を見ながら採用するか、或いはスキップするかを選択する（採用する場合は特に何もする必要はなく、スキップする場合だけ、早送りを選択する）。

【0043】ユーザが映像を見終わると、最新の映像重要度に対応する数値データの重み付けを計算する。

【0044】採用された映像に含まれる 数値データ ( $X_i$ ) の値を  $k_i$  としたとき、 $X_i$  が  $k_i$  を平均、 $\sigma^2$  を分散とする正規分布の値をパラメータ・ポテンシャルに投票する。

【0045】次に他の時区間で数値データ ( $X_i$ ) の値が  $h_i$  のとき、同様に投票値が加算され、数値データ ( $X_i$ ) に対するパラメータ・ポテンシャルは図2のようになる。

【0046】ダイジェスト映像における加算回数  $S$  と  $X_i$  におけるパラメータ・ポテンシャルの総和が0となる閾値  $Th$  を求め、パラメータ・ポテンシャルを0~1に正規化する。

【0047】他の数値データ ( $X_n$ ) についても同様に、パラメータ・ポテンシャルを計算する。

【0048】加算回数  $S$  が小さいとき、パラメータ・ポテンシャルの値はばらつくため、閾値  $Th$  は大きな値を持つが、 $S$  が大きくかつ選択されるシーンとパラメータに関連性が大きくなると、パラメータ・ポテンシャルは先鋭化し、そのとき閾値  $Th$  は小さくなる。

【0049】後述するが、重要度を決めるときポテンシャル値と閾値の差の総和を求めるため、先鋭化したポテンシャル値が強調されることになる。

【0050】また、数値データを複数用いる場合は、複数の数値データを相互に関連付けるため、2次元ポテンシャル・グラフを生成する。

【0051】同一フレームにおける数値データ ( $X_i$ ) と ( $X_j$ ) の値がそれぞれ  $k_i$ 、 $p_j$  のとき、平均が  $k_i$ 、 $p_j$  で分散が  $\sigma^2$  の2次元正規分布を加算する。

【0052】ポテンシャルの総和が0になるように閾値  $Th_2$  を求め、2次元ポテンシャル・グラフを0~1に正規化する。 $n$  種類の数値データに対して  $n(n-1)/2$  の2次元ポテンシャル・グラフが生成できる。

【0053】一方、ユーザが映像をスキップしたとき、その映像はそれ以降で積極的に採用されることがないよう、強化学習の手段を用いて、パラメータポテンシャルを変更する。

【0054】ここで強化学習とは、一般に外部システムの判断による成功に対し報酬を与え学習を進める手法のことである。

【0055】スキップされた映像に含まれる 数値データ ( $X_i$ ) の値を  $k_i$  としたとき、 $X_i$  が  $k_i$  を平均、 $\sigma$  を分散とし、 $k_i$  のポテンシャル  $p(k_i)$  に対し  $p(k_i)/2$  倍の正規分布を減ずる。

【0056】他の数値データ ( $X_n$ ) 及び、複数の数値データによる2次元ポテンシャル・グラフについても同様に減算して、ポテンシャル・グラフを変更する。

【0057】次に、パラメータ・ポテンシャル及び2次元ポテンシャルグラフを利用して重要度を定める。

【0058】各フレームにおける映像の数値データの (ポテンシャル値- $Th$ ) 及び2次元ポテンシャル・グ

ラフの (ポテンシャル値- $Th_2$ ) の総和を求め、シグモイド関数により重要度を決定する。

【0059】シグモイド関数における  $T=c/S$  ( $c$ : 定数) として加算回数によって重要度の確信度を向上させる。

【0060】以上のようにパラメータ・ポテンシャルを所有しているユーザは、次回からサンプルの重要度を使わずにポテンシャル・グラフから個人の映像重要度を設定することができ、更に、映像を見ながら「採用」或いは「スキップ」を選択することで、ポテンシャル・グラフが変更され、ユーザの好みが強反映される。

【0061】また、ユーザは、ポテンシャル・グラフから求めた重要度を、セマンティック・データを用いて変更できる。

【0062】セマンティック・データの注釈に一致するキーワードをユーザが入力すると、該当する時区間の重要度を最大にし、重要度に基づく要約が行われる。

【0063】このときの入力状況もポテンシャル・グラフに反映する。

【0064】以上が本発明の処理の流れの概要であり、以下で図を用いて更に詳細に説明する。

【0065】図3は、本発明の全体の流れを示したフローチャートである。

【0066】ステップ110において、本発明の処理が開始される。

【0067】ステップ120において、重要度の算出が行われる。

【0068】ステップ130において、画面の表示及びユーザの選択処理が行われる。

【0069】ステップ140において、パラメータ・ポテンシャルの更新を行うか否かがチェックされ、そのチェックの結果、パラメータ・ポテンシャルの更新を行う場合は、ステップ170に移行し、パラメータ・ポテンシャルの更新を行わない場合は、ステップ150へ移行する。

【0070】ステップ170において、強化学習法を用いたパラメータ・ポテンシャルの更新が行われる。

【0071】このパラメータ・ポテンシャルを更新は、このコンテンツの終了時、又はユーザが更新を望んだときに行われる。

【0072】ステップ180において、パラメータ・ポテンシャルの選択を行う。

【0073】ユーザのパラメータ・ポテンシャルが存在しないときは、デフォルト値 (例えば、コンテンツ配信者が定める) を用い、ユーザのパラメータ・ポテンシャルが存在する場合はそれを使う。

【0074】ステップ150において、全フレームが終了したかチェックされる。

【0075】例えば、1つのコンテンツ全てを見終わったとき終了とすることができる。

10

20

30

40

50

【0076】ステップ160において、次フレームを取り出し、またステップ120へ移行する。

【0077】ステップ190において、処理を終了する。

【0078】図4は、図3の重要度算出処理（ステップ120）を更に詳細に説明したフローチャートである。

【0079】ステップ210において、重要度算出処理が開始される。

【0080】ステップ220において、現フレームにおける数値データ $X_i$  ( $i=1, 2, \dots, n$ ) において\*10

$$sum\_p = \sum_{i=1}^n (p(X_i) - Th\_i) + \sum_{i,j} (p(X_i, X_j) - Th2\_ij)$$

【0084】ステップ250において、シグモイド関数により重要度を求める。

【0085】ここで、シグモイド関数は以下の式で与えられる。

【0086】

【数2】

$$P(X) = 1 - \frac{1}{\exp(sum\_p/T)}$$

ただし、 $T=c/S$  ( $c$ :定数)、 $S$ は投票データ数とする。

【0087】図5は、図3の画像表示及びユーザの選択・投票処理（ステップ130）を更に詳細に説明したフロー・チャートである。

【0088】ステップ310において、画像表示及びユーザの選択・投票処理（ステップ130）が開始される。

【0089】ステップ320において、要約を行うか否かが判断され、要約を行う場合は、ステップ330に進み、要約を行わない場合は、ステップ340に進む。

【0090】ステップ330において、重要度が所定の閾値 $Th\_Imp$ 未満のときは、このシーンは重要でないと考えられ、スキップされステップ370へ進む。

【0091】一方、重要度が所定の閾値 $Th\_Imp$ 以上のときは、このシーンはスキップされず、ステップ340へ進む。

【0092】ステップ340において、現フレームの画面が表示される。

【0093】ステップ350において、現画面を見ているユーザが、このシーンを見たい(Accept)か、見たくない(Reject)か、或いは、どちらもでもない(neutral)かの選択(判断)を行う。

【0094】この判断の結果、見たい(Accept)又は見たくない(Reject)の場合は、ステップ360に進み、どちらもでもない(neutral)の場合は、ステップ370へ進む。

\*それぞれ $p(X_i) - Th\_i$ の値を計算する。

【0081】ステップ230において、現フレームにおける数値データの組 $(X_i, X_j)$  ( $1 \leq i, j < n$ , となる全ての組み合わせ)においてそれぞれ $p(X_i, X_j) - Th2\_ij$ の値を計算する。

【0082】ステップ240において、ステップ220及び230で得られた全ての値の和(sum\_p)を求める。

【0083】

【数1】

【0095】ステップ360において、現フレームの数値データ $X_i$  ( $i=1, 2, \dots, n$ ) を見たい(Accept)又は見たくない(Reject)の結果に応じて投票し、記憶する。

【0096】図6は、図3の強化学習によるパラメータ・ポテンシャル更新処理（ステップ170）を更に詳細に説明したフローチャートである。

【0097】ステップ410において、強化学習によるパラメータ・ポテンシャル更新処理が開始される。

【0098】ステップ420において、投票データ(数値データ $X_i$  ( $i=1, 2, \dots, n$ ))を取り出す。

【0099】ステップ430において、取り出された投票データが見たい(Accept)又は見たくない(Reject)か判断され、見たい(Accept)の場合は、ステップ440へ進み、見たくない(Reject)の場合は、ステップ450へ進む。

【0100】ステップ440において、 $X_i$ の値が $k_i$ とすると、平均 $k_i$ 、分散 $\sigma^2$ とする正規分布の値を $X_i$ のパラメータ・ポテンシャルに加算し、 $i=1 \sim n$ まで同じ処理を繰り返す。

【0101】同時に、 $(X_i, X_j)$ の値がそれぞれ $(k_i, p_j)$ とすると、平均 $(k_i, p_j)$ 、分散 $\sigma^2$ の2次元正規分布を $(X_i, X_j)$ のポテンシャル・グラフに加算し、全ての $(i, j)$  ( $i=1 \sim n, j=1 \sim m$ )の組に対して、同じ処理を繰り返す。

【0102】ステップ450において、 $X_i$ の値が $k_i$ とすると、平均 $k_i$ 、分散 $\sigma$ とし、 $k_i$ のポテンシャル $p(k_i)$ に対して $p(k_i)/2$ の正規分布を $X_i$ のパラメータ・ポテンシャルから減じ、 $i=1 \sim n$ まで同じ処理を繰り返す。

【0103】同時に、 $(X_i, X_j)$ の値がそれぞれ $(k_i, p_j)$ とすると、平均 $(k_i, p_j)$ 、分散 $\sigma$ とし、 $(k_i, p_j)$ のポテンシャル $p(k_i, p_j)$ に対して、 $p(k_i, p_j)/2$ の2次元正規分布を $(X_i, X_j)$ のポテンシャル・グラフから減じ、全て



の $(i, j)$ の組に対して、同じ処理を繰り返す。

【0104】ステップ460において、得られた全てのパラメータ・ポテンシャルに対して、それぞれ総和が0になるように閾値 $Th1_i$ を求め、0～1に正規化する。

【0105】同様に2次元ポテンシャル・グラフに対しても、閾値 $Th2_j$ を求め、0～1に正規化する。

【0106】ステップ470において、強化学習によるパラメータ・ポテンシャル更新処理が終了する。

【0107】図7には、本発明に実施に適したハードウェアの一例としてPVR (personal Video Recorder) 10の主な構成がブロックで示されている。

【0108】ブロック510は、ビデオ蓄積部であり、長時間のビデオ・データを蓄積可能なハードディスク装置やビデオ・テープ・レコーダー等の大容量記憶装置が利用される。

【0109】ブロック520は、本発明を適用したソフトウェアである。

【0110】ブロック530は、ビデオ・バッファであり、重要度に基づいて要約されたビデオを一時的に保存し、ハードディスク装置等の記憶装置が用いられる。 20

【0111】ビデオデータ505は、PVRのビデオ蓄積部510に入力され、保存される。

【0112】ビデオ蓄積部510は、入力されたビデオ・データ505を順次記憶する。

【0113】ビデオ蓄積部510に記憶されたビデオ・データは、本発明を適用したソフトウェア520により処理され、重要度に応じて要約が作成される。

【0114】要約されたビデオ・データは、ビデオ・バッファ530に一時的に記憶されユーザが希望するとき 30にビデオ出力535として出力される。

【0115】一方、ビデオ蓄積部に記憶されたビデオ・データ515は、要約されずにそのままビデオ出力として出力することも可能である。

【0116】図8～図10には、これまで説明した本発明を実際のサッカーの映像への適用例が示されている。

【0117】本発明のユーザ・インタフェースとしては、図8に示したようなウィンドウ800が考えられる。

【0118】ウィンドウ800は、映像表示領域850、学習用の領域810、要約用の領域820等を含む。 40

【0119】「Start」(スタート)ボタン804をクリックすることにより対象の映像が映像表示領域850に表示が開始され、「Stop」(ストップ)ボタン806をクリックすることにより表示中の映像が停止され、「Pause」(ポーズ)ボタン808をクリックすると表示中の映像が一時停止する。

【0120】まず、学習方法は、以下のような処理を行う。

【0121】画面を見ながら、ユーザはその画面が重要視したい画像ならば「Accept」ボタン812をクリックし、重要視しない画像ならば「Reject」ボタン814をクリックする。

【0122】また、「Update」(更新)816ボタンをクリックすることで、パラメータ・ポテンシャルが更新され、ユーザの好みが反映される。

【0123】パラメータ・ポテンシャル自体をユーザは見ることができないが、各画面の重要度818の数値を見ることにより、ユーザの好みが適切に反映されているかを確認することができる。

【0124】次に、要約方法は、以下のような処理を行う。

【0125】ラジオ・ボタンを「On」(オン)822にすることにより、現フレームから最終フレームにかけて重要度が算出され、その重要度の算出結果に基づいて対象となる映像の要約が行われる。

【0126】表示領域830は要約結果を示し、白い部分はスキップするシーンを表し、黒い部分は要約するシーンを表している。

【0127】すなわち、要約を表示する場合には、黒い部分のシーンのみ表示されることになる。

【0128】更に、要約表示中もユーザは「Accept」ボタン812又は「Reject」ボタン814を適宜クリックすることにより、更に好みを反映することもできる。

【0129】要約用の領域820には、閾値826や要約率(%)828を表示することも可能である。

【0130】閾値826は、要約とし採用するシーンの最低の重要度を示している。

【0131】要約率(%)とは、対象映像の表示時間と要約映像の表示時間の比をパーセントで表したものである。

【0132】図9には、図8に示したサッカーの適用例に使用する数値データの実例が示されている。

【0133】図8のようなサッカーの試合を表す数値データとしては、カメラモーション(カメラの向き)、領域分割結果、平均色情報、及びこれらの数値データのパラメータ相関等が用いられる。

【0134】カメラモーションとしては、パン(左右の向き(角度))、チルト(上下の向き(角度))の2種類の値が用いられる。

【0135】図9にカメラモーションのパン(左右の向き)についてのパラメータ・ポテンシャル910が示されている。

【0136】横軸912は、パンの向き(角度)を表し、左側が角度-90度で右側が角度+90度である。

【0137】縦軸914は、各パンの角度に対応して数値を表している。

50 【0138】パラメータ・ポテンシャル910におい

て、1点破線920は、投票前のパラメータ・ポテンシャルを表し、破線918はユーザによる投票データ、実線916は投票後のパラメータ・ポテンシャルを表している。

【0139】領域分割結果としては、最大面積領域の面積(S)、最大面積領域の色(r-g、b-y)の値の3種類の値が用いられる。

【0140】図9に領域分割結果の最大面積領域の面積(S)についてのパラメータ・ポテンシャル930が示されている。

【0141】横軸932は、最大面積領域の面積(S)を表し、左側が面積0で右側に行くほど面積が大きくなる。

【0142】縦軸934は、各最大面積領域の面積(S)に対応して数値を表している。

【0143】パラメータ・ポテンシャル930において、1点破線940は、投票前のパラメータ・ポテンシャルを表し、破線938はユーザによる投票データ、実線936は投票後のパラメータ・ポテンシャルを表している。

【0144】平均色情報として、R(赤)、G(緑)、B(青)の3種類の値が用いられる。

【0145】図9に平均色情報G(緑)についてのパラメータ・ポテンシャル950が示されている。

【0146】横軸952は、平均色情報G(緑)を表し、左側が面積0で右側に行くほどGが大きくなる。

【0147】縦軸954は、各平均色情報G(緑)に対応して数値を表している。

【0148】パラメータ・ポテンシャル950において、1点破線960は、投票前のパラメータ・ポテンシャルを表し、破線958はユーザによる投票データ、実線956は投票後のパラメータ・ポテンシャルを表している。

【0149】また、更に上記の合計8つのパラメータの相関値28種類の値も利用される。

【0150】従って、図8に示されたサッカーの例では、合計36種類の数値データが重要度を決定するために使用される。

【0151】しかし、重要度を計算するために使用する数値データの種類の、対象となる映像により異なり、また、数値データの種別を変えることにより要約の精度も異なってくる。

【0152】カメラモーション(カメラの向き)、領域分割結果、平均色情報の数値データは、原データを使う場合と、データを主成分分析した特徴データを使うことも可能である。

【0153】しかし、原データをそのまま使うと、データ間の相関を無視しているため、寄与するデータとそうでないデータが同じ尺度で扱われ、相関の高いデータでは、2重投票を行ったのと同じ効果になる場合があると

いう問題が生じる。

【0154】従って、これらの原データに対し、主成分分析を行い、データ間の相関が小さくなる特徴量に置き換えた投票を行うことにする。

【0155】ここで、主成分分析とは、n次元のデータを互いに独立な少数個の指標で表そうとすることである。

【0156】主成分分析では、n次のパラメータ群に対し、相関のもっとも小さい順に並べたm次(m<n)の特徴量で表現する。

【0157】図9に示した各データは、原データを使った場合の投票の様子を表したものである。

【0158】図10には、図9を用いて説明して数値データからもとめた映像の各シーンの重要度が示されている。

【0159】図10の横軸1002は、時間を表し、縦軸1004は、重要度が示されている。

【0160】いま希望する要約を作成するために適した重要度の閾値が1006で示されている。

【0161】すなわち、対象となるサッカー映像のうち閾値以上の重要度を有する部分は、要約に含められ(シーンA、B、C、D)、閾値未満の重要度を有する部分(斜線の部分)は、要約には含まれずカットされることになる。

【0162】以上により、図10に示された例の場合には、時間が約半分に短縮された要約が作成される。

【0163】

【効果】本願発明の構成によって、ユーザの好みに沿った映像の要約を作成することが可能となる。

【0164】また、映像を見ながら更にユーザの好みを要約に反映させることも可能となる。

【0165】以下まとめとして他の実施例を記載する。

【0166】(1)映像から要約を作成する方法であって、対象となる映像から数値データを抽出するステップと、前記抽出した数値データに基づいて各シーンの重要度を決定するステップと、前記重要度が所定の閾値以上の部分から要約を作成するステップと、を含む方法。

(2)映像から要約を作成する方法であって、対象となる映像から数値データを抽出するステップと、前記対象となる映像がセマンティック・データを含む場合には当該セマンティック・データ及び前記数値データに基づいて各シーンの重要度を決定し、前記対象となる映像がセマンティック・データを含まない場合には、前記数値データに基づいて各シーンの重要度を決定するステップと、前記重要度が所定の閾値以上の部分から要約を作成するステップと、を含む方法。

(3)映像から要約を作成する方法であって、対象となる映像から複数の種類の数値データを抽出するステップと、前記抽出した複数の種類の数値データに基づいて各シーンの重要度を決定するステップと、前記重要度が所

定の閾値以上の部分から要約を作成するステップと、を含む方法。

(4) 映像から要約を作成する方法であって、対象となる映像から数値データを抽出するステップと、対象となる映像を表示中にユーザの好みを表すデータを受け取るステップと、前記抽出した数値データ及び前記ユーザの好みを表すデータに基づいて各シーンの重要度を決定するステップと、前記重要度が所定の閾値以上の部分から要約を作成するステップと、を含む方法。

(5) 映像から要約を作成する方法であって、対象となる映像から数値データを抽出するステップと、前記抽出した数値データに基づいて各シーンの重要度を決定するステップと、前記重要度が所定の閾値以上の部分から要約を作成するステップと、前記作成された要約を表示するステップと、前記要約を表示中にユーザの入力により前記重要度を変更するステップと、を含む方法。

(6) 映像から要約を作成する装置であって、対象となる映像から数値データを抽出する手段と、前記抽出した数値データに基づいて各シーンの重要度を決定する手段と、前記重要度が所定の閾値以上の部分から要約を作成する手段と、を含む装置。

(7) 映像から要約を作成する装置であって、対象となる映像から数値データを抽出する手段と、前記対象となる映像がセマンティック・データを含む場合には当該セマンティック・データ及び前記数値データに基づいて各シーンの重要度を決定し、前記対象となる映像がセマンティック・データを含まない場合には、前記数値データに基づいて各シーンの重要度を決定する手段と、前記重要度が所定の閾値以上の部分から要約を作成する手段と、を含む方法。

(8) 映像から要約を作成する装置であって、対象となる映像から複数の種類の数値データを抽出する手段と、前記抽出した複数の種類の数値データに基づいて各シーンの重要度を決定する手段と、前記重要度が所定の閾値以上の部分から要約を作成する手段と、を含む装置。

(9) 映像から要約を作成する装置であって、対象となる映像から数値データを抽出する手段と、対象となる映像を表示中にユーザの好みを表すデータを受け取る手段と、前記抽出した数値データ及び前記ユーザの好みを表すデータに基づいて各シーンの重要度を決定する手段と、前記重要度が所定の閾値以上の部分から要約を作成する手段と、を含む装置。

(10) 映像から要約を作成する装置であって、対象となる映像から数値データを抽出する手段と、前記抽出した数値データに基づいて各シーンの重要度を決定する手段と、前記重要度が所定の閾値以上の部分から要約を作

成する手段と、前記作成された要約を表示する手段と、前記要約を表示中にユーザの入力により前記重要度を変更する手段と、を含む装置。

(11) 映像から要約を作成するシステムであって、前記システムは、プロセッサと、メモリと、入力された映像を記憶する記憶装置と、映像要約装置とを含み、前記映像要約装置は、対象となる映像から数値データを抽出する手段と、前記抽出した数値データに基づいて各シーンの重要度を決定する手段と、前記重要度が所定の閾値以上の部分から要約を作成する手段と、を含むシステム。

(12) 映像から要約を作成するプログラムを記憶する記録媒体であって、前記プログラムは、対象となる映像から数値データを抽出するステップと、前記抽出した数値データに基づいて各シーンの重要度を決定するステップと、前記重要度が所定の閾値以上の部分から要約を作成するステップと、を含む記録媒体。

【図面の簡単な説明】

【図1】 サンプル映像の重要度とシーンの関係を示した図である。

【図2】 パラメータポテンシャルへ有効パラメータの投票を示した図である。

【図3】 本発明の全体の処理を示すフローチャートである。

【図4】 本発明の重要度算出処理を示すフローチャートである。

【図5】 本発明の画面表示、ユーザの選択処理を示すフローチャートである。

【図6】 本発明の強化学習によるパラメータポテンシャル更新処理を示すフローチャートである。

【図7】 本発明を適用したPVRの構成を示す図である。

【図8】 サッカー映像への本発明の適用例を示す図である。

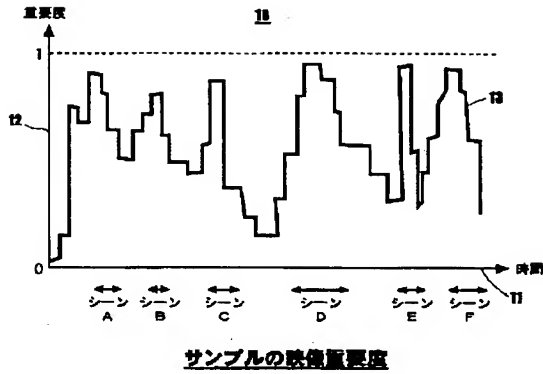
【図9】 本発明に使用される数値データの例を示す図である。

【図10】 本発明を適用した場合のシーンと要約の関係を示す図である。

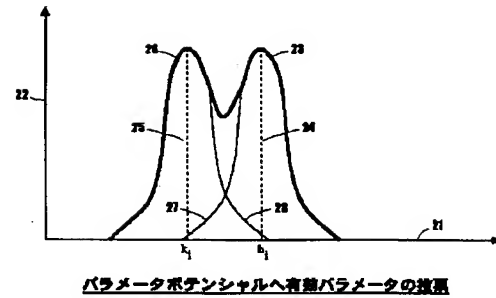
【符号の説明】

500 PVR  
505 ビデオ入力  
510 ビデオ蓄積部  
520 映像要約部  
530 ビデオ・バッファ  
535 ビデオ出力

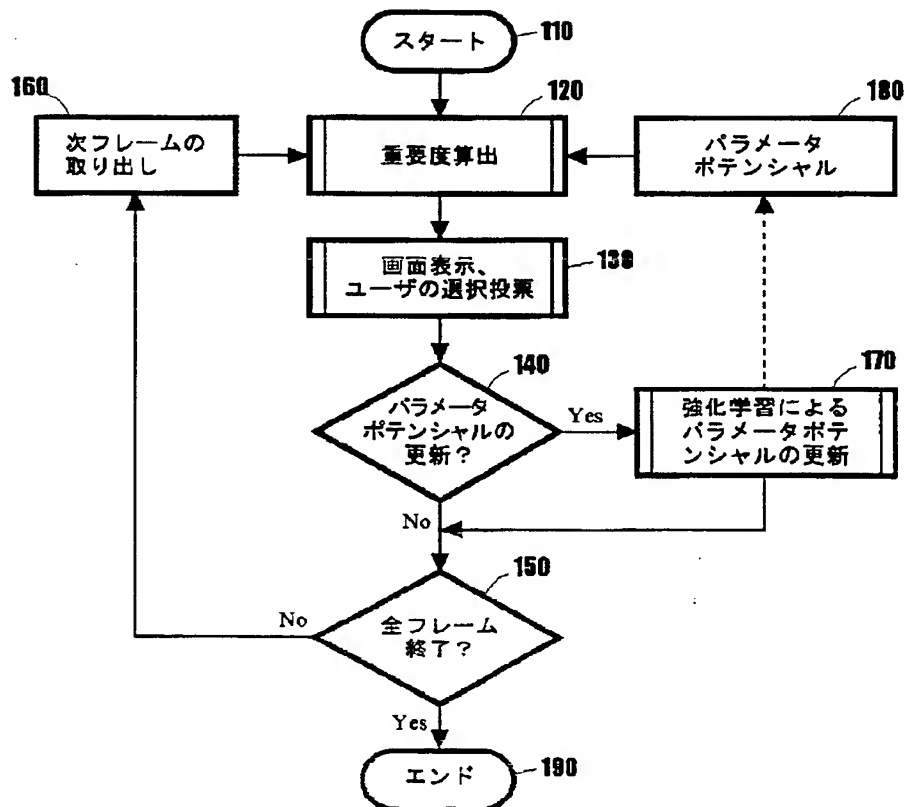
【図1】



【図2】

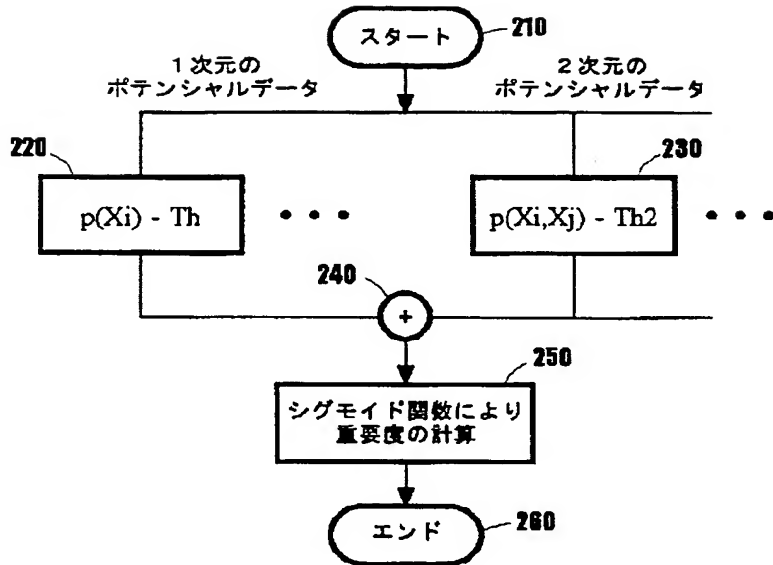


【図3】

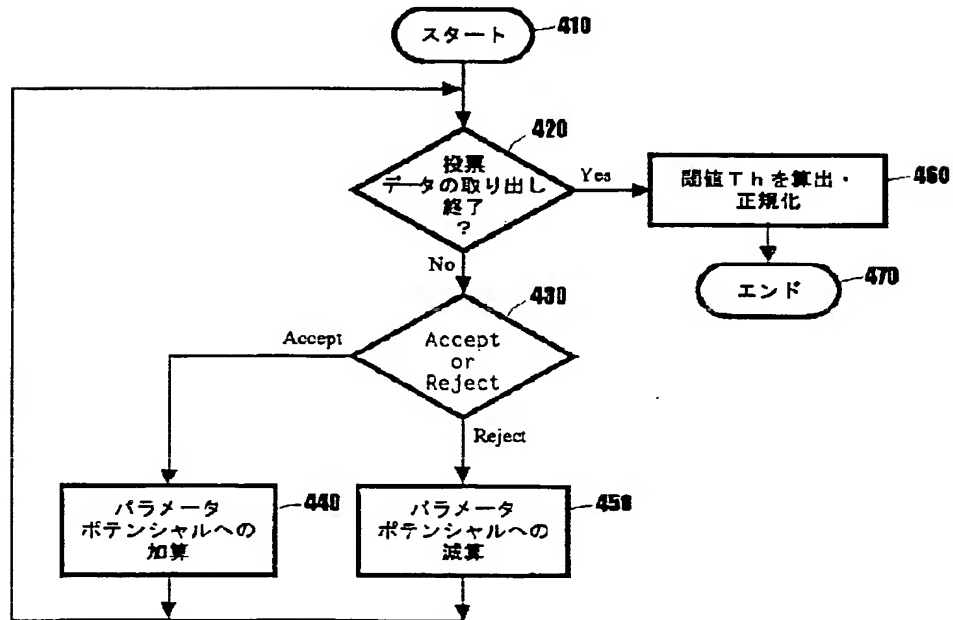


全体のフローチャート

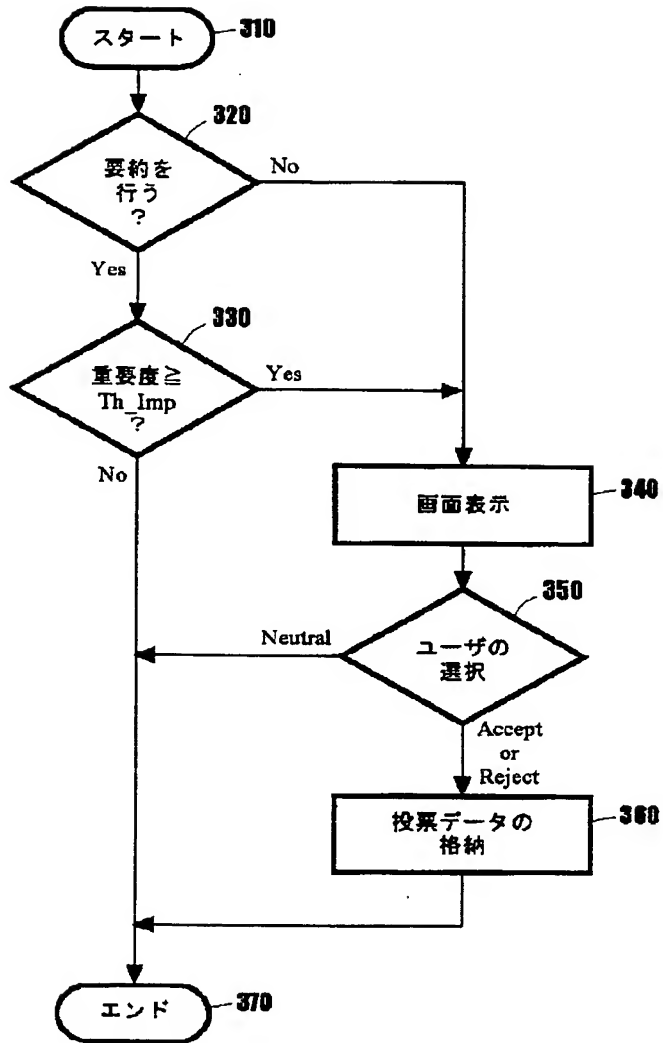
【図4】

重要度算出のフローチャート

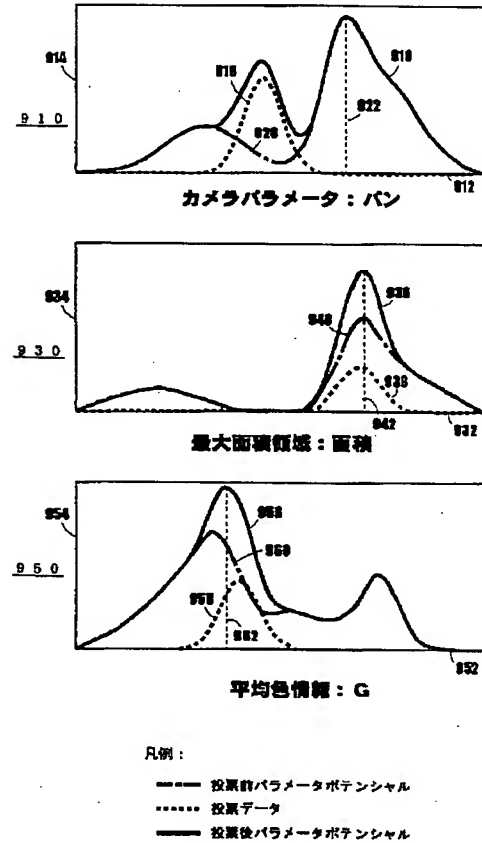
【図6】

強化学習によるパラメータポテンシャル更新のフローチャート

【図5】

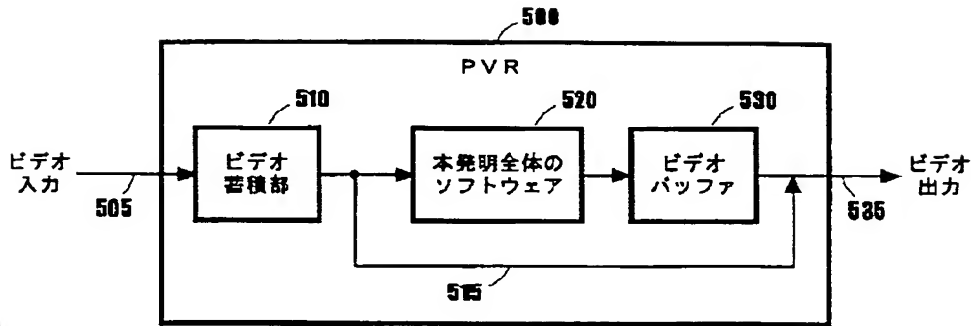


【図9】

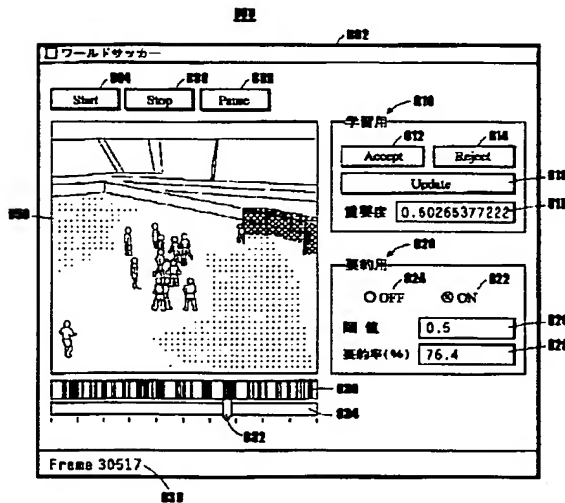


画面表示、ユーザの選択のフローチャート

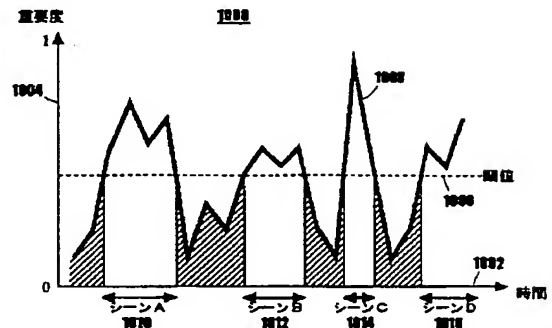
【図7】

**PVRの構成**

【図8】

**スポーツ(サッカー)映像における再描写**

【図10】

**重要度から要約シーンの求め方**

フロントページの続き

(72)発明者 越後 富夫  
 神奈川県大和市下鶴間1623番地14 日本ア  
 イ・ビー・エム株式会社 東京基礎研究所  
 内

(72)発明者 富田 アルベルト  
 神奈川県大和市下鶴間1623番地14 日本ア  
 イ・ビー・エム株式会社 東京基礎研究所  
 内

(72)発明者 益満 健  
 神奈川県大和市下鶴間1623番地14 日本ア  
 イ・ビー・エム株式会社 東京基礎研究所  
 内

Fターム(参考) 5B075 ND12 NS01 PQ02 PR03  
 5C052 AA01 AB03 AB04 AC08 DD04  
 EE02 EE03  
 5C053 FA23 GB21 HA29 JA21 KA04  
 KA05 KA24 LA06 LA07

THIS PAGE BLANK (USPTO)



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record.**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☒ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)